DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 10/31/07 has been entered.

This communication is responsive to Amendment filed 10/31/07, 03/05/08.

Claims 1-39 are pending in this application. Claims 1, 31, 32, 33, 38 are independent claims. This action is made non-Final.

Specification

2. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required:

Claims 33-37 recite the limitation "a computer readable medium", however, the term "a computer readable medium" is not found in the Specification. There is insufficient antecedent basis for this limitation.

Claim Objections

3. Claims 33-37 are objected to because of the following informalities: Claims 32-37 recite the limitations "A computer-readable medium…". The term "a computer-readable medium" is

not supported by the specification. There are insufficient antecedent basis for these limitations in the claim. Applicant will be required to make appropriate amendment to the description to provide clear support or antecedent basis for the terms appearing in the claims provided no new matter is introduced.

Claim Rejections - 35 USC § 101

4. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

5. Claims 33-39 are rejected under 35 U.S.C. § 101 because the claimed invention is directed to non-statutory subject matter.

(a) Regarding claim 33, Applicant has amended claim 33 as "A computer- program product, residing on a computer readable medium", however, the claimed program product is not limited to embodiments, which include the hardware necessary to enable any underlying functionality to be realized. As such, the medium is reasonably interpreted as just software, thus, the claim lacks the necessary physical articles or objects to constitute a machine or a manufacture within the meaning of 35 U.S.C. § 101. It is recommended that the term "computer readable medium" should be read as "computer readable **storage** medium", and this term should also be supported by the specification.

Claims 34-38 are dependent upon claim 18, respectively, suffer from deficiencies similar to their respective base claims, and therefore are likewise rejected.

(b) Claim 38 recites "a computer system", however, each recited element of the system may be reasonably interpreted by one of ordinary skill as software modules; for example,

par [0204] indicates "*This system 510, in turn, is comprised of four subsystems, a subsystem for similarity search 512, a subsystem for information retrieval 514, a subsystem for clustering 516, and a subsystem for matching 518. As described above, similarity search may rely on the inverted indexes of the information retrieval subsystem*" suggesting the claim as a whole can be implemented using software means only, as these subsystems that make up the system are all software applications that do not result in a tangible practical application under 35 U.S.C. § 101; thus, the system is not tangible embodied in a manner so as to be executable. The claim lacks the necessary physical articles or objects to constitute a machine or a manufacture within the meaning of 35 U.S.C. § 101, instead being software per se. See MPEP 2106.01.

As such, the claimed system does not define any specific hardware and needs to be amended to include physical computer hardware (e.g. processor, memory) to execute the software components.

Claim Rejections - 35 USC § 103

6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time

a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

7. Claims 1-23, 26-36, 38, 39 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madan et al. (US Patent No. 6,778,980), in view of Jannink et al. (US Patent No. 6,697,800), and further in view of Gallivan et al. (US Patent No. 6,778,995).

As per claim 1, Madan teaches a computer-implemented method for searching a collection of items *(i.e. to a set of electronic commerce pages, category pages and databases for locating products electronically, col. 2, lines 26-48)*, wherein each item in the collection has a set of properties *(i.e. words, col. 2, lines 26-48)*, comprising the steps of:

obtaining a query *(i.e. The search string, col. 4, lines 63-67)* composed of a first set of one or more properties *(i.e. Acme, Mint, Mouthwash, col. 4, lines 63-67)*; and

obtaining a result based on applying a distance function to the query and an item *(i.e. hit vector, col. 5, line 65 to col. 6, line 5)* in the collection having a second set of one or more properties *(i.e. A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5)* wherein

obtaining a result includes determining a third set of properties *(i.e. Acme, Mint, col. 6, lines 30-40)* common to the first set of one or more properties and the second set of one or more properties *(i.e. A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5)*, and

the distance function determines a distance between the query and the item in the collection based on a frequency of occurrence of items *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* in the collection that have all of the properties in the third set of properties, wherein a higher frequency of occurrence of items *(i.e. Acme: 48 matches; Mint: 168 matches, col. 6,*

lines 30-40) in the collection having all of the properties in the third set of properties indicates a greater distance between the query and the item in the collection and a lower frequency of occurrence of items *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* in the collection having all of the properties in the third set of properties indicates a smaller distance between the query and the item in the collection *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)*; and

providing a representation of the result to a user *(i.e. Search results are provided at 430, for example, by displaying all search results simultaneously. For example, looking at the top hit vector, 0, 0, 0, 0, 0, 1,1, 1, 0 (group 1) in the Acne Mint Mouthwash example, only two bits are set to "1." This means that all of the search terms in the search string did not match. Since only the third and second bits from the right are set to "1," all the matches in group 1 got there because they contained the first and second search terms in the search string "Acme Mint Mouthwash." Thus, the sub-string result set is displayed, for example: Acme Mint (18 matches), col. 6, lines 54-64).*

Note that Madan implicitly teaches "distance function *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the*

potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).

However, Jannink fairly teaches this limitation *(i.e. affinity represents a measure of the distance, or similarity, between two items. Affinity between two items can be calculated as the measure of the difference between the items' normalized properties. Those items that are close, i.e., have a relatively small distance between them, are considered to have a stronger affinity than those items that are further apart. FIG. 4 is a flowchart that describes operation 304 in greater detail according to embodiment of the present invention, illustrating the preferred approach to computing an affinity between a search item and other items in the objective database 104, col. 5, line 59 to col. 6, line 2).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan and Jannink at the time the invention was made to modify the system of Madan to include the limitations as taught by Jannink. One of ordinary skill in the art would be motivated to make this combination in order to measure the distance between two items in view of Jannink (col. 5, line 59 to col. 6, line 2), as doing so would give the added benefit of providing an improved system and method for determining affinity between items of data using both objective and subjective data as taught by Jannink (col. 1, lines 47-49).

Further note that Madan implicitly teaches "frequency occurrence" *(i.e. Acme: 48 matches; Mint: 168 matches; Acme AND Mint: 18 matches, col. 6, lines 30-40).*

But Gallivan clearly teaches this limitation *(i.e. Histograms of the frequency of occurrences of the terms and phrases in each document and over the entire document set are generated. Related documents are identified by finding highly correlated term and phrase*

pairings, Summary).

It would have been obvious to one of ordinary skill of the art having the teaching of Madan, Jannink and Gallivan at the time the invention was made to modify the system of Madan, Jannink to include the limitations as taught by Gallivan. One of ordinary skill in the art would be motivated to make this combination in order to build a multi-dimensional semantic concept space over a stored document collection in view of Gallivan (Summary), as doing so would give the added benefit of achieving an approach to forming clusters of concepts and themes into groupings of classes with shared semantic meanings. Such an approach would preferably categorize concepts mined from a document set into clusters defined within a pre-specified range of variance. Moreover, such an approach would not require a priori knowledge of the data content as taught by Gallivan (col. 2, lines 36-44).

As per claim 31, Madan teaches a computer-implemented method for analyzing two sets of properties form a plurality of sets of properties of items in a collection being search, each item in the collection having a set of properties in the plurality of sets of properties, the method comprising the steps of:

determining a set properties *(i.e. Acme, Mint, col. 6, lines 30-40)* common to the two sets of properties *(i.e. Acme, Mint, Mouthwash, col. 4, lines 63-67; A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5)*;

determining a frequency of occurrence of sets properties, in the plurality of sets of properties, that include all the common properties *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)*;

assessing the distance between the two sets of properties as a function of the frequency of occurrence of sets of properties that include all the common properties, wherein a higher frequency of occurrence *(i.e. Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40)* of sets of properties that include all the common properties indicates a greater distance, and a lower frequency of occurrence *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* of sets of properties that include all the common properties indicates a smaller distance *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)*; and

providing a representation of the distance to a user *(i.e. Search results are provided at 430, for example, by displaying all search results simultaneously. For example, looking at the top hit vector, 0, 0, 0, 0, 0, 1,1, 1, 0 (group 1) in the Acne Mint Mouthwash example, only two bits are set to "1." This means that all of the search terms in the search string did not match. Since only the third and second bits from the right are set to "1," all the matches in group 1 got there because they contained the first and second search terms in the search string "Acme Mint Mouthwash." Thus, the sub-string result set is displayed, for example: Acme Mint (18 matches), col. 6, lines 54-64).*

Although Madan implicitly teaches "distance function" *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a*

large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50), Jannink fairly teaches this limitation *(i.e. affinity represents a measure of the distance, or similarity, between two items. Affinity between two items can be calculated as the measure of the difference between the items' normalized properties. Those items that are close, i.e., have a relatively small distance between them, are considered to have a stronger affinity than those items that are further apart. FIG. 4 is a flowchart that describes operation 304 in greater detail according to embodiment of the present invention, illustrating the preferred approach to computing an affinity between a search item and other items in the objective database 104, col. 5, line 59 to col. 6, line 2).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan and Jannink at the time the invention was made to modify the system of Madan to include the limitations as taught by Jannink. One of ordinary skill in the art would be motivated to make this combination in order to measure the distance between two items in view of Jannink (col. 5, line 59 to col. 6, line 2), as doing so would give the added benefit of providing an improved system and method for determining affinity between items of data using both objective and subjective data as taught by Jannink (col. 1, lines 47-49).

Furthermore, Madan also implicitly suggests "frequency occurrence" *(i.e. Acme: 48 matches; Mint: 168 matches; Acme AND Mint: 18 matches, col. 6, lines 30-40).* But, Gallivan fairly teaches this limitation *(i.e. Histograms of the frequency of occurrences of the terms and phrases in each document and over the entire document set are generated. Related documents are identified by finding highly correlated term and phrase pairings, Summary).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan, Jannink and Gallivan at the time the invention was made to modify the system of Madan, Jannink to include the limitations as taught by Gallivan. One of ordinary skill in the art would be motivated to make this combination in order to build a multi-dimensional semantic concept space over a stored document collection in view of Gallivan (Summary), as doing so would give the added benefit of achieving an approach to forming clusters of concepts and themes into groupings of classes with shared semantic meanings. Such an approach would preferably categorize concepts mined from a document set into clusters defined within a pre-specified range of variance. Moreover, such an approach would not require a priori knowledge of the data content as taught by Gallivan (col. 2, lines 36-44).

As per claim 32, Madan teaches a computer-implemented method for analyzing the relationship between two items in a collection of items, wherein each item in the collections is associated with a set of properties, comprising the steps of:

obtaining a set of properties *(i.e. Acme, Mint, col. 6, lines 30-40)* with which the two items are commonly associated *(i.e. Acme, Mint, Mouthwash, col. 4, lines 63-67; A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5)*; and

determining the degree of commonality between the two items as a function of a frequency of occurrence of items in the collection that have all of the properties which the two items have in common, wherein a higher frequency occurrence *(i.e. Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40)* of items in the collection having all of the properties which the

two items have in common indicates a lesser degree of commonality and a lower frequency of occurrence *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* of items in the collection having all of the properties which the two items have in common indicates a greater degree of commonality *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)*; and

providing a representation of the degree of commonality to a user *(i.e. Search results are provided at 430, for example, by displaying all search results simultaneously. For example, looking at the top hit vector, 0, 0, 0, 0, 0, 1,1, 1, 0 (group 1) in the Acne Mint Mouthwash example, only two bits are set to "1." This means that all of the search terms in the search string did not match. Since only the third and second bits from the right are set to "1," all the matches in group 1 got there because they contained the first and second search terms in the search string "Acme Mint Mouthwash." Thus, the sub-string result set is displayed, for example: Acme Mint (18 matches), col. 6, lines 54-64).*

Madan implicitly teaches "distance function" *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).*

Jannink, in addition, fairly teaches this limitation *(i.e. affinity represents a measure of the distance, or similarity, between two items. Affinity between two items can be calculated as the measure of the difference between the items' normalized properties. Those items that are close, i.e., have a relatively small distance between them, are considered to have a stronger affinity than those items that are further apart. FIG. 4 is a flowchart that describes operation 304 in greater detail according to embodiment of the present invention, illustrating the preferred approach to computing an affinity between a search item and other items in the objective database 104, col. 5, line 59 to col. 6, line 2).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan and Jannink at the time the invention was made to modify the system of Madan to include the limitations as taught by Jannink. One of ordinary skill in the art would be motivated to make this combination in order to measure the distance between two items in view of Jannink (col. 5, line 59 to col. 6, line 2), as doing so would give the added benefit of providing an improved system and method for determining affinity between items of data using both objective and subjective data as taught by Jannink (col. 1, lines 47-49).

Madan implicitly teaches "frequency occurrence" *(i.e. Acme: 48 matches; Mint: 168 matches; Acme AND Mint: 18 matches, col. 6, lines 30-40).*

Gallivan explicitly teaches this limitation *(i.e. Histograms of the frequency of occurrences of the terms and phrases in each document and over the entire document set are generated. Related documents are identified by finding highly correlated term and phrase pairings, Summary).*

It would have been obvious to one of ordinary skill of the art having the teaching of

Madan, Jannink and Gallivan at the time the invention was made to modify the system of Madan, Jannink to include the limitations as taught by Gallivan. One of ordinary skill in the art would be motivated to make this combination in order to build a multi-dimensional semantic concept space over a stored document collection in view of Gallivan (Summary), as doing so would give the added benefit achieving an approach to forming clusters of concepts and themes into groupings of classes with shared semantic meanings. Such an approach would preferably categorize concepts mined from a document set into clusters defined within a pre-specified range of variance. Moreover, such an approach would not require a priori knowledge of the data content as taught by Gallivan (col. 2, lines 36-44).

As per claim 33, Madan teaches a computer program product, residing on a computer readable medium, for use in searching a collection of items, wherein each item in the collection is associated with a set of properties, the computer program product comprising instructions for causing a computer to:

receive a query *(i.e. The search string, col. 4, lines 63-67)* composed of one or more properties *(i.e. Acme, Mint, Mouthwash, col. 4, lines 63-67)*; and

obtain a result based on applying a distance function to the query and an item *(i.e. hit vector, col. 5, line 65 to col. 6, line 5)* in the collection having a second set of one or more properties *(i.e. A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5)*;

wherein the distance function determines a third set of properties *(i.e. Acme, Mint, col. 6, lines 30-40)* common to the first set of one or more properties and the second set of one or more

properties, and determines a distance between the query and the item in the collection based on a frequency of occurrence of items in the collection that have all of the properties in the third set of properties, wherein a higher frequency of occurrence *(i.e. Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40)* of items in the collection having all of the properties in the third set of properties indicates a greater distance between the query and the item in the collection and a lower frequency of occurrence *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* of items in the collection having all of the properties in the third set of properties indicates a smaller distance between the query and the item in the collection *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)*; and

provide a representation of the result to a user *(i.e. Search results are provided at 430, for example, by displaying all search results simultaneously. For example, looking at the top hit vector, 0, 0, 0, 0, 0, 1,1, 1, 0 (group 1) in the Acne Mint Mouthwash example, only two bits are set to "1." This means that all of the search terms in the search string did not match. Since only the third and second bits from the right are set to "1," all the matches in group 1 got there because they contained the first and second search terms in the search string "Acme Mint Mouthwash." Thus, the sub-string result set is displayed, for example: Acme Mint (18 matches), col. 6, lines 54-64).*

Madan implicitly teaches "distance function" *(i.e. To compensate for the zero hits*

resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).

In complement, Jannink teaches this limitation *(i.e. affinity represents a measure of the distance, or similarity, between two items. Affinity between two items can be calculated as the measure of the difference between the items' normalized properties. Those items that are close, i.e., have a relatively small distance between them, are considered to have a stronger affinity than those items that are further apart. FIG. 4 is a flowchart that describes operation 304 in greater detail according to embodiment of the present invention, illustrating the preferred approach to computing an affinity between a search item and other items in the objective database 104, col. 5, line 59 to col. 6, line 2).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan and Jannink at the time the invention was made to modify the system of Madan to include the limitations as taught by Jannink. One of ordinary skill in the art would be motivated to make this combination in order to measure the distance between two items in view of Jannink (col. 5, line 59 to col. 6, line 2), as doing so would give the added benefit of providing an improved system and method for determining affinity between items of data using both objective and subjective data as taught by Jannink (col. 1, lines 47-49).

Madan suggests "frequency occurrence" *(i.e. Acme: 48 matches; Mint: 168 matches; Acme AND Mint: 18 matches, col. 6, lines 30-40).*

Gallivan teaches this limitation *(i.e. Histograms of the frequency of occurrences of the terms and phrases in each document and over the entire document set are generated. Related documents are identified by finding highly correlated term and phrase pairings, Summary).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan, Jannink and Gallivan at the time the invention was made to modify the system of Madan, Jannink to include the limitations as taught by Gallivan. One of ordinary skill in the art would be motivated to make this combination in order to build a multi-dimensional semantic concept space over a stored document collection in view of Gallivan (Summary), as doing so would give the added benefit of achieving an approach to forming clusters of concepts and themes into groupings of classes with shared semantic meanings. Such an approach would preferably categorize concepts mined from a document set into clusters defined within a pre-specified range of variance. Moreover, such an approach would not require a priori knowledge of the data content as taught by Gallivan (col. 2, lines 36-44).

As per claim 38, Madan teaches a computer system for managing data records comprising:

an information retrieval subsystem that stores and retrieves data records *(i.e. to a set of electronic commerce pages, category pages and databases for locating products electronically, col. 2, lines 26-48)*, each data record being associated with a set of properties *(i.e. words, col. 2, lines 26-48)*; and

a similarity search subsystem that receives similarity search queries *(i.e. The search string, col. 4, lines 63-67)* and processes similarity search queries based on a distance function

(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50), a similarity search query being associated with a first set of properties *(i.e. Acme, Mint, Mouthwash, col. 4, lines 63-67)*;

wherein the distance function determines a distance between the query and a data record in the collection having a second set of properties *(i.e. Acme, Mint, col. 6, lines 30-40)* based on determining the frequency of occurrence of data records in the collection that are associated with all of the properties in the third set of properties *(i.e. Acme, Mint, col. 6, lines 30-40)*, wherein a higher frequency of occurrence *(i.e. Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40)* of data records associated with all of the properties in the third set of properties indicates a greater distance between the query and the item and a lower frequency of occurrence *(i.e. Acme AND Mint: 18 matches, col. 6, lines 30-40)* of data records associated with all of the properties in the third set of properties indicates a smaller distance between the query and the item.

Although the terms "distance function" and "similarity search" are implied in the teaching of Madan *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)*, Jannink fairly teaches this limitation *(i.e. As described above, affinity represents a measure of the*

distance, or similarity, between two items. Affinity between two items can be calculated as the measure of the difference between the items' normalized properties. Those items that are close, i.e., have a relatively small distance between them, are considered to have a stronger affinity than those items that are further apart. FIG. 4 is a flowchart that describes operation 304 in greater detail according to embodiment of the present invention, illustrating the preferred approach to computing an affinity between a search item and other items in the objective database 104, col. 5, line 59 to col. 6, line 2).

It would have been obvious to one of ordinary skill of the art having the teaching of Madan and Jannink at the time the invention was made to modify the system of Madan to include the limitations as taught by Jannink. One of ordinary skill in the art would be motivated to make this combination in order to measure the distance between two items in view of Jannink (col. 5, line 59 to col. 6, line 2), as doing so would give the added benefit of providing an improved system and method for determining affinity between items of data using both objective and subjective data as taught by Jannink (col. 1, lines 47-49).

As per claim 2, the step of associating each item in the collection with a set of properties *(i.e. Consider the following illustrative example. The objective database 104 includes four items (A1, A2, A3, and A4) each item having three properties (x, y, z), col. 6, lines 52-55).*

As per claim 3, the step of obtaining a result includes identifying one or more result items whose distance form the query is within a first threshold *(i.e. In addition, grouping a number of objective items into an appropriate one or more clusters, wherein the number of*

objective items grouped have the affinity value within a predetermined distance with respect to the search item, Summary).

As per claim 4, Jannink teaches the step of obtaining a result includes ranking the one or more result items according to their distance for the query *(i.e. Items A2, A3, and A4 can then be ranked in order of their affinity, from smallest (the greatest affinity to search item A1) to largest (the least affinity to search item A1). Assume for purposes of illustration that item A4 has the smallest affinity value, followed by A2 and then A3 with the highest affinity value. The initial ranking is therefore A4, A2, A3, where affinity is calculated using objective data stored in objective database 104, col. 7, lines 23-30).*

As per claim 5, Madan teaches the threshold is defined as a number of result items *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).*

As per claim 6, Jannink teaches the threshold is define as a distance *(i.e. In addition, grouping a number of objective items into an appropriate one or more clusters, wherein the number of objective items grouped have the affinity value within a predetermined distance with respect to the search item, Summary).*

As per claim 7, Madan teaches the step of returning the result *(i.e. Search results are provided at 430, for example, by displaying all search results simultaneously. For example, looking at the top hit vector, 0, 0, 0, 0, 0, 1,1, 1, 0 (group 1) in the Acne Mint Mouthwash example, only two bits are set to "1." This means that all of the search terms in the search string did not match. Since only the third and second bits from the right are set to "1," all the matches in group 1 got there because they contained the first and second search terms in the search string "Acme Mint Mouthwash." Thus, the sub-string result set is displayed, for example: Acme Mint (18 matches), col. 6, lines 54-64).*

As per claim 8, Madan teaches the step of obtaining a query includes the step of mapping a received query to a set of one or more properties *(i.e. A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5).*

As per claim 9, Madan teaches one or more properties are binary *(i.e. A hit vector of "0, 0, 0, 0, 0, 1, 1, 0" indicates a match of the search terms "Acme" and "Mint,", col. 5, line 65 to col. 6, line 5).*

As per claim 10, Madan teaches one or more properties are related by a partial order, and wherein, it an item is associated with a property, then the item is also associates with all ancestors of that property in the partial order *(i.e. Acme AND Mint: 18 matches, Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40).*

As per claim 11, Madan teaches one or more of the properties represent numerical values or ranges, and wherein the partial order reflects a set of containment relationship among the numerical values or ranges *(i.e. Acme AND Mint: 18 matches, Acme: 48 matches; Mint: 168 matches, col. 6, lines 30-40)*.

As per claim 12, Jannink teaches the properties are group into equivalence class *(i.e. A function may be defined for each class of items that can be used to provide the appropriate canonical label, given a wide variety of typical variations of the label that are often used to refer to the item. In this way, users 112 who misspell or use a shortened version of a label to refer to a particular item are mapped to the correct search item (i.e., the item intended by the user), col. 5, lines 39-49)*.

As per claim 13, the step of grouping the properties into equivalence classes using clustering *(i.e. Updating the one or more clusters, wherein the plurality of objective items are rearranged between the one or more clusters as a result of the adjusted affinity value, Summary)*.

As per claim 14, Jannink teaches each properties has a set of sub-properties, wherein the clustering is performed such that the distance between two properties in the collection is correlated to the number of properties in the collection that are associated with all of the sub-properties common to both properties *(i.e. A method of determining a relationship between a search item provided by a user and a plurality of objective items. Each objective item in the plurality includes a plurality of objective item properties and the search item includes one or*

more search item properties. The method comprises grouping a subset of objective items based on an objective relationship between one or more of the plurality of objective item properties and the one or more search item properties, Summary).

As per claim 15, Jannink teaches the query corresponds to a single item in the collection *(i.e. A method of determining a relationship between a search item provided by a user and a plurality of objective items. Each objective item in the plurality includes a plurality of objective item properties and the search item includes one or more search item properties. The method comprises grouping a subset of objective items based on an objective relationship between one or more of the plurality of objective item properties and the one or more search item properties, Summary).*

As per claim 16, Jannink teaches the query corresponds to a plurality of items in the collection *(i.e. In operation 408, the normalized differences are combined to form an affinity measurement between the search item and the target item. These steps are then repeated to generate an affinity measurement for each target item in objective database 104, col. 6, lines 34-51).*

As per claim 17, Madan teaches the query is independent of the items in the collection *(i.e. The search string, Acme, Mint, Mouthwash, col. 4, lines 63-67).*

As per claim 18, Jannink teaches the step of obtaining a result is constrained to a sub-collection of the items in the collection *(i.e. A method of determining a relationship between a search item provided by a user and a plurality of objective items. Each objective item in the plurality includes a plurality of objective item properties and the search item includes one or more search item properties. The method comprises grouping a subset of objective items based on an objective relationship between one or more of the plurality of objective item properties and the one or more search item properties, Summary).*

As per claim 19, Jannink teaches the sub-collection is specified as an expression of properties *(i.e. A function may be defined for each class of items that can be used to provide the appropriate canonical label, given a wide variety of typical variations of the label that are often used to refer to the item. In this way, users 112 who misspell or use a shortened version of a label to refer to a particular item are mapped to the correct search item (i.e., the item intended by the user), col. 5, lines 39-49).*

As per claim 20, Jannink teaches the expression includes a subset of the set of properties of the set of properties that compose the query *(i.e. A function may be defined for each class of items that can be used to provide the appropriate canonical label, given a wide variety of typical variations of the label that are often used to refer to the item. In this way, users 112 who misspell or use a shortened version of a label to refer to a particular item are mapped to the correct search item (i.e., the item intended by the user), col. 5, lines 39-49).*

As per claim 21, Madan teaches the step of obtaining a query includes identifying certain properties to be ignored in the step of obtaining a result *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).*

As per claim 22, Jannink teaches the distance function is applied explicitly *(See formula in col. 8, line 67).*

As per claim 23, Jannink teaches the distance function is applied implicitly *(See formula in col. 7, lines 1-5).*

As per claim 26, Madan teaches the step of obtaining a result includes iterating through one or more subsets of the query and identifying items associated with the one or more subsets *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50)..*

As per claim 27, Madan teaches the one or more subsets are prioritized according to the number of items in the collection that have all the properties each subset and wherein iterating through one or more subsets of the query is continued until a first threshold is reached *(i.e. To compensate for the zero hits resulting from the AND search, the electronic commerce Web site can provide or perform an OR search. If the search performed is an OR search, (e.g., Acme OR Mint OR Mouthwash), a large number of hits may result. The number of hits can be so large as to overwhelm the potential customer, which can result in the loss of potential sales. What is needed is an improved search engine, col. 1, lines 44-50).*

As per claim 28, Jannink teaches the step of obtaining a result includes applying a Euclidean distance function *(i.e. Euclidean, col. 17, lines 1-22).*

As per claim 29, Jannink teaches the step of obtaining a result includes merging a first result determined by applying the distance function and a second result determined by applying the Euclidean distance function *(See formulas in col. 7, lines 1-5 and col. 8, line 67).*

As per claim 30, Jannink teaches the step of obtaining a result determined by applying the Euclidean distance function *(i.e. Euclidean, col. 17, lines 1-22).*

As per claim 34, Jannink teaches the instruction cause the computer to obtain a result by identifying exactly the items whose distance form the query is within a threshold *(i.e. In addition, grouping a number of objective items into an appropriate one or more clusters,*

wherein the number of objective items grouped have the affinity value within a predetermined distance with respect to the search item, Summary).

As per claim 35, Jannink teaches the instruction cause the computer to obtain a result by identifying approximately the items whose distance from the query is within a threshold according to a heuristic *(i.e. Also, computing a subjective property value for each rule, wherein the subjective property indicates a strength of the user-defined relationship, col. 2, lines 1-25).*

As per claim 36, Jannink teaches the heuristic permits a trade-off between the accuracy and the performance of a search *(i.e. The length of the time window can therefore be varied to trade-off statistical reliability of response time with changing user opinion, col. 9, line 42 to col. 10, line 7).*

As per claim 39, Gallivan teaches a clustering subsystem that employs the distance function of the similarity search subsystem to construct a graph *(See Figs. 9, 14).*

8. Claims 24, 25, 37 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madan et al. (US Patent No. 6,778,980), in view of Jannink et al. (US Patent No. 6,697,800), and in view of Gallivan et al. (US Patent No. 6,778,995), and further in view of Antoshenkov et al (US Patent No. 5,379,422).

As per claim 24, Madan, Jannink, Gallivan do not specifically teach the step of obtaining a result includes the step of iterating a random walk process to select potential result items.

Antoshenkov teaches this limitation *(i.e. leaf node selection means for selecting a leaf node by a random walk down the tree from the root node to a leaf node, wherein said means for selecting includes acceptance/rejection selection means for selecting a child node of a previously selected intermediate node with a selection probability based upon the approximate ranking information of the child node and the approximate ranking information of the previously selected intermediate node, col. 106, lines 36-45).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan, Jannink, Gallivan, Antoshenkov at the time the invention was made to modify the system of Madan, Jannink, Gallivan to include the limitations as taught by Antoshenkov. One of ordinary skill in the art would be motivated to make this combination in order to select a leaf node at random by descending from the root node and selecting at each node a child of the node by a random selection weighted by the cardinality estimates of the children in view of Antoshenkov (col. 3, lines 40-59), as doing so would give the added benefit of increasing the deviation reduces database maintenance at the expense of increased rejections, and decreasing the deviation reduces the rejections at the expense of increased maintenance as taught by Antoshenkov (col. 4, lines 26-34).

As per claim 25, Antoshenkov teaches ranking the potential result items *(i.e. The preferred method of FIG. 12 can be seen to give simple and unbiased sampling by another proof based on the similarity of the preferred method to the method of random sampling on ranked trees, col. 16, line 58 to col. 17, line 19).*

Gallivan teaches the step of obtaining a result includes ranking the potential result items by frequency and selecting the potential items having higher frequency *(i.e. Finally, the most highly correlated terms and phrases from the extracted concepts are categorized (transition 65) into clusters 49, col. 6, lines 3-10).*

As per claim 37, Madan, Jannink, Gallivan do not specifically teach the heuristic includes the use of a random walk process.

However, Antoshenkov teaches this limitation *(i.e. leaf node selection means for selecting a leaf node by a random walk down the tree from the root node to a leaf node, wherein said means for selecting includes acceptance/rejection selection means for selecting a child node of a previously selected intermediate node with a selection probability based upon the approximate ranking information of the child node and the approximate ranking information of the previously selected intermediate node, col. 106, lines 36-45).*

It would have been obvious to one of ordinary skill of the art having the teaching of Madan, Jannink, Gallivan, Antoshenkov at the time the invention was made to modify the system of Madan, Jannink, Gallivan to include the limitations as taught by Antoshenkov. One of ordinary skill in the art would be motivated to make this combination in order to select a leaf node at random by descending from the root node and selecting at each node a child of the node by a random selection weighted by the cardinality estimates of the children in view of Antoshenkov (col. 3, lines 40-59), as doing so would give the added benefit of increasing the deviation reduces database maintenance at the expense of increased rejections, and decreasing

the deviation reduces the rejections at the expense of increased maintenance as taught by Antoshenkov (col. 4, lines 26-34).

Response to Arguments

9. Applicant's arguments with respect to claims 1-39 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Miranda Le whose telephone number is (571) 272-4112. The examiner can normally be reached on Monday through Friday from 8:30 AM to 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John R. Cottingham, can be reached on (571) 272-7079. The fax number to this Art Unit is (571)-273-8300.

Any inquiry of a general nature or relating to the status of this application should be directed to the Group receptionist whose telephone number is (571) 272-2100.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/Miranda Le/

Primary Examiner, Art Unit 2167